Operational Pipeline for Large-scale 3D Reconstruction of Buildings from Satellite Images

Sebastien Tripodi, Liuyun Duan, Veronique Poujade, Frederic Trastour, Jean-Philippe Bauchet, Lionel Laurore, Yuliya Tarabalka

OPERATIONAL PIPELINE FOR LARGE-SCALE 3D RECONSTRUCTION OF BUILDINGS FROM SATELLITE IMAGES

Sebastien Tripodi, Liuyun Duan, Veronique Poujade, Frederic Trastour, Jean-Philippe Bauchet, Lionel Laurore, Yuliya Tarabalka

LuxCarta Technology, 06370 Mouans-Sartoux, France. Email: stripodi@luxcarta.com

ABSTRACT

Automatic 3D reconstruction of urban scenes from stereo pairs of satellite images remains a popular yet challenging research topic, driven by numerous applications such as telecommunications and defense. The quality of reconstruction results depends particularly on the quality of the available stereo pair. In this paper, we propose an operational pipeline for large-scale 3D reconstruction of buildings from stereo satellite images. The proposed chain uses U-net to extract contour polygons of buildings, and the combination of optimization and computational geometry techniques to reconstruct a digital terrain model and a digital height model, and to correctly estimate the position of building footprints. The pipeline has proven to be efficient for 3D building reconstruction, even if the close-to-nadir image is not available.

Index Terms— 3D reconstruction, satellite images, building footprint, deep learning, digital surface model.

1. INTRODUCTION

A few recent years have witnessed an increasing interest in the topic of 3D reconstruction of urban scenes from stereo satellite images. While until recently the quality of satellite imagery coupled with existing methodologies did not allow to produce 3D city models at a high-spatial resolution in an automatic way [1], very-high-resolution commercial satellites (Worldview, Pleiades) launched in the last decade acquire high-quality stereo images all over the Earth, with a spatial resolution of up to 30 cm/pixel. This boosted the development of stereo reconstruction methods in remote sensing community.

One of the first methods for urban scene reconstruction in LOD1 (model where buildings have flat roofs) has used a semi-global matching (SGM) technique [2] to find correspondences in a stereo pair of epipolar images, followed by a joint classification using image radiometry coupled with estimated elevation information to retrieve 3D city models [3]. Even though this method offered a solution for 3D urban reconstruction at a large scale, small geometries could not be captured precisely. The recently released benchmarks for large-scale semantic 3D reconstruction [4, 5] further intensified the research on this topic. The winning solutions of the 2019 IEEE GRSS data fusion challenge mostly used U-net or ResNet for semantic labeling, and SGM or pyramid stereo matching network for disparity estimation [4].

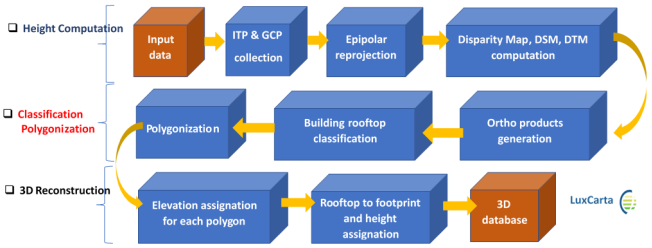


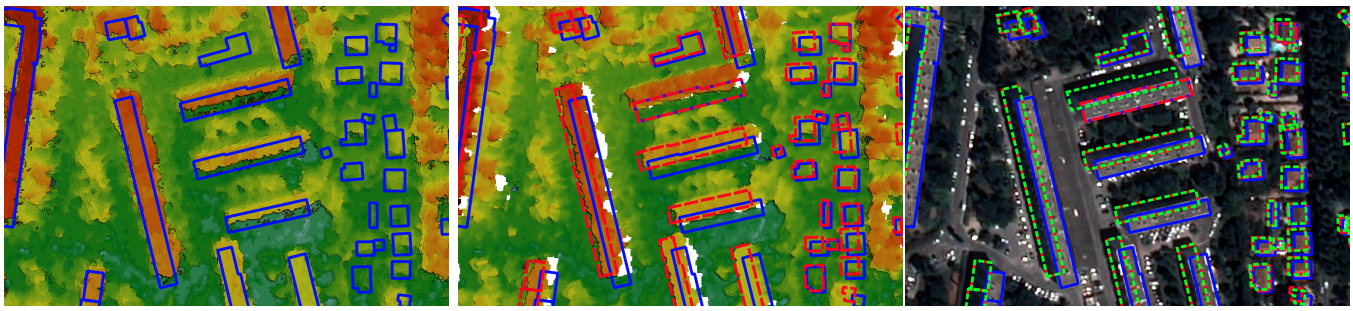
Fig. 1: Proposed pipeline.

In most existing works on semantic 3D reconstruction from a stereo pair of satellite images [6, 7], the output is a disparity map together with semantic segmentation of one or both given images. In this paper, we propose a complete operational pipeline, which takes as an input one stereo pair of satellite images and the corresponding rational polynomial coefficient (RPC) models [8], and automatically reconstructs a 3D model, consisting of a digital terrain model (DTM) and vectors of building footprints together with their heights in LOD1. One of important contributions is a method which automatically projects building rooftops extracted by U-net from a single image to footprints (bases of buildings in the geographical coordinate system).

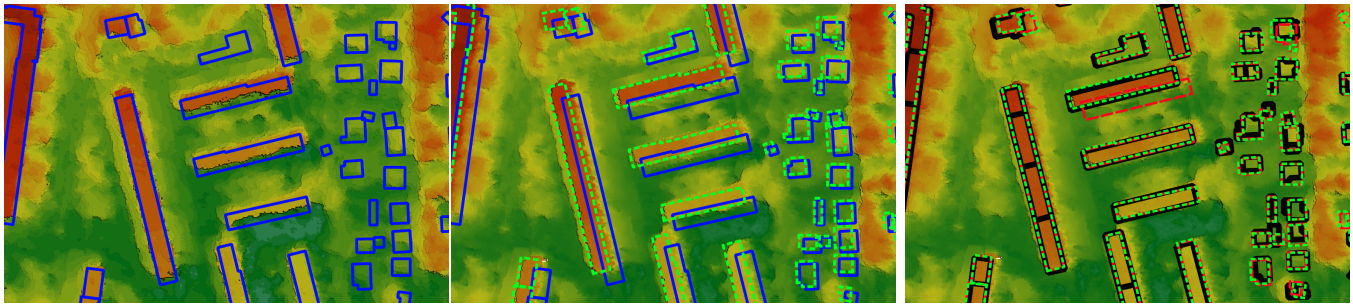
2. PROPOSED PIPELINE

The proposed chain for large-scale 3D reconstruction of urban scenes in LOD1 is described in Fig. 1. The input is a stereo pair of high-resolution satellite images (acquired by different satellites with different spatial resolutions, such as Worldview, Pléiades, GeoEye, Spot), with the associated RPC models provided by the vendors. In this article, we apply our pipeline at spatial resolution of 50 cm/pixel, which is enough for extracting the rooftop contours and footprints. The outputs are a DTM and a set of building 3D models in LOD1, $i.e$ a set of polygons with the associated height for each building.

Our chain consists of tree main parts presented in the fol-



(a) Rooftop on DSM$_{sgm}$ in fake ortho space (b) Rooftop vs footprint on DSM$_{sgm}$ (c) Rooftop vs footprint on fake ortho image



(d) Rooftop on DSM$_{lux}$ in fake ortho space (e) Rooftop vs footprint on DSM$_{lux}$ (f) Footprint vs ground truth on DSM$_{lux}$

Fig. 2: Results of the rooftop to footprint algorithm. Rooftop, footprint using DSM$_{lux}$, footprint using DSM$_{sgm}$, ground truth.

lowing sub-sections:

- Height computation by computing a Digital Surface Model (DSM) and a DTM.
- Semantic labeling by extracting the building rooftop.
- 3D reconstruction in LOD1, $i.e$ shifting rooftop polygons to their corresponding footprints and assigning a height to each buildings.

2.1. Height computation

From the input stereo pair of images (level 2A), height information is extracted by using an epipolar geometry to compute a disparity map and a DSM. The first step is to adjust the RCP model (we call the resulting model *RPC-adjust*) for georeferencing the images and computing epipolar images. We refer the reader to [9] for details of the designed algorithm, based on the AKAZE feature detection.

Reconstructing an accurate DSM and DTM is a critical part of the chain, because the elevation information is further used for extracting a height of every building but also for shifting contours of rooftops to their corresponding footprints. The algorithm must be robust to manage very different geographic areas: flat/mountainous region, different kinds of buildings, $etc.$. In addition, if for a given area of interest the available satellite images are limited, the algorithm has to manage a fake stereo pair, $i.e$ different vintage, different satellite sensors and cloudy regions. As shown in [4], two main approaches exist to compute a disparity map between two stereo images: based on SGM and deep learning. Due to the difficulty to build a ground-truth representing different scenarios mentioned above, we retained the method based on SGM. We modified the original SGM algorithm [2], to enable solving certain conditions such as textureless regions. The improvements proposed in this paper include:

1. *Pyramidal approach*. The SGM is executed at different resolution scales from 8 to 1. We thus remove noise by estimating consistency between resolutions.
2. *Census* as a cost function to be more robust to radiometric difference, due for ex. to shadows or clouds.
3. GPU implementation for fast execution. It also manages a large disparity range by splitting this range to fit the GPU memory, executing the SGM for each new range and merging these results.

The DSM is computed from the disparity map and RPC model. Fig. 2b and 2e show the improvements obtained by our algorithm (DSM$_{lux}$) compared to the standard SGM (DSM$_{sgm}$): contours are sharper, there are significantly less noisy and non-informed values. Quantitative evaluation of the DSM is difficult, because it is too laborious to manually build a DSM ground-truth. Using other sources (ex., LIDAR) to evaluate DSM is not accurate due to vintage and data misalignment. We propose to evaluate the DSM quality by using this DSM to move the building rooftop polygons to their corresponding footprints, and compare the outputs with the manually drawn footprints (see Sec. 3). A DTM is further computed using our DTM generation algorithm [9] consisting of two steps: classification and surface interpolation. This DTM is used to orthorectify the closest-to-nadir image.

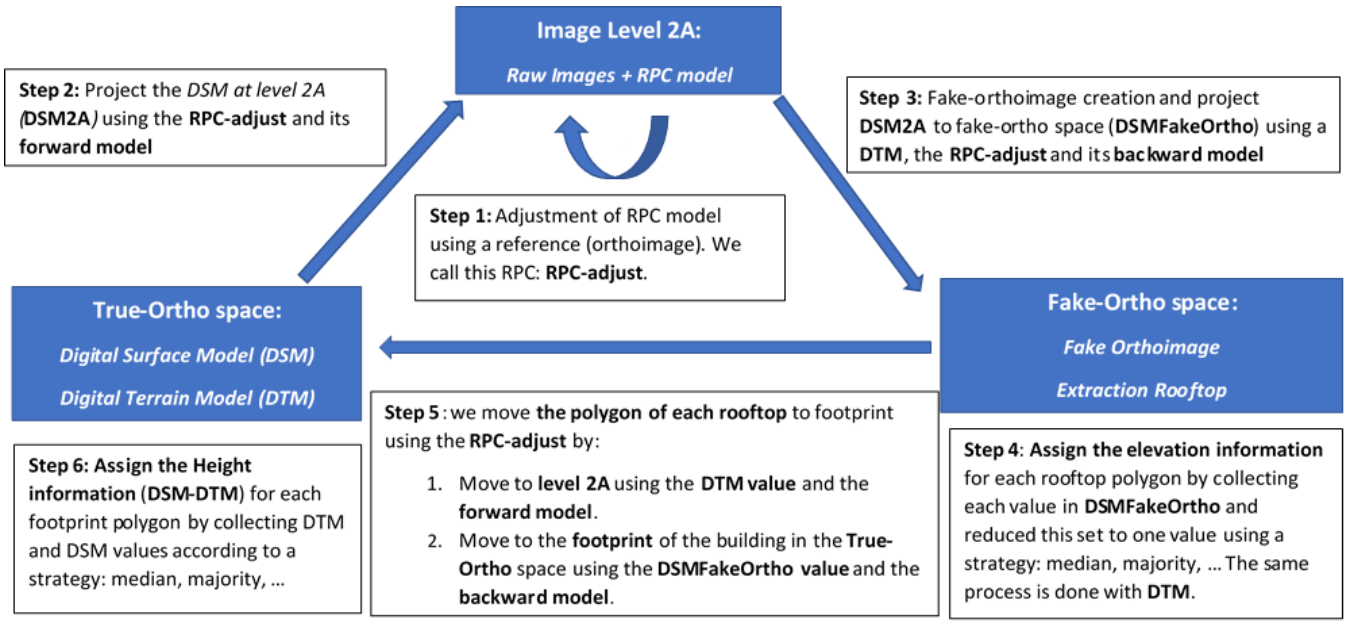


Fig. 3: Proposed rooftop to footprint projection algorithm.

2.2. Rooftop extraction

The algorithm for extracting rooftop polygons consists of two steps: 1) Semantic labeling of an orthoimage. 2) Polygonization of building contours.

Semantic labeling means classifying each pixel of the image to a *building* or *no-building* class. For this task, we have adopted a U-net convolutional neural network architecture, which has exhibited the highest performance in several benchmarks involving building rooftop labeling [10, 4]. A detailed description of the designed model and the Luxcarta database used for its training in given in [11]. We have shown in [11] that a careful design of the model together with an appropriate loss function (we have used a combination of cross-entropy and intersection over union losses) allows to train a generic model, which performs well on different types of urban areas, such as residential, industrial and very dense areas.

While U-net outputs classification in a raster format, these results must be further polygonized, to yield a vector of each building rooftop. We have designed a solution to polygonize building contours, which performs a naive polygonization of the mask of every building, followed by a polygon simplification, which searches for a compressed polygon with the best quality/complexity ratio, *i.e.* with the minimum number of vertices within a specified tolerance of an error [11].

2.3. 3D reconstruction

The last part of our chain aims at reconstructing buildings in 3D-LOD1, $i.e$ assigning a unique height value for every building and moving a rooftop to the footprint. Both problems of the height assignation and 'rooftop to footprint' are linked, as we explain here-below. We apply the commonly used approach to create an orthoimage, by rectifying the initial level 2A image with the DTM. We call this rectification as 'fake orthoimage.' The interest to use the DTM for orthorectification is to avoid getting non-informed values in occluded areas. The inconvenience is that for any off-nadir image (majority of use cases) building rooftops in the resulting orthoimage do not coincide with the corresponding footprints (see Fig. 2c). This implies a shift between the DSM, which is in the true orthospace, and rooftop polygons (see Fig. 2e), making the assignment of building heights problematic.

To avoid this problem, a DSM could be used to produce a true orthoimage. However, the DSM computed from satellite images or LIDAR deforms geometry of objects (e.g., rooftops), and the resulting orthoimage would not be suitable for building contour detection. In this paper, we generalize the problem and propose a solution, which shifts rooftop polygons to footprints, using an RPC model and any DSM and DTM. This solution is geometrically exact; another advantage is that if a DSM is provided from another source (e.g., LIDAR), we do not need a stereo pair of images, only one off-nadir image is sufficient to reconstruct a 3D model.

Fig. 3 explains each step of the proposed method. The main idea is to project the DSM in the fake orthoimage space (see Fig. 2a and 2d). In this space the elevation values of the DSM coincide correctly with rooftops. These values are collected within each rooftop polygon, and a building elevation is estimated by retaining the median or the majority value. Knowing the elevation of each building, we move rooftop polygons to the corresponding footprints, which coincide both in fake and true orthospace. The footprints are thus well aligned with the DSM in the true orthospace; this allows to assign a height for every building by substracting DTM from DSM and retaining the median or majority value,

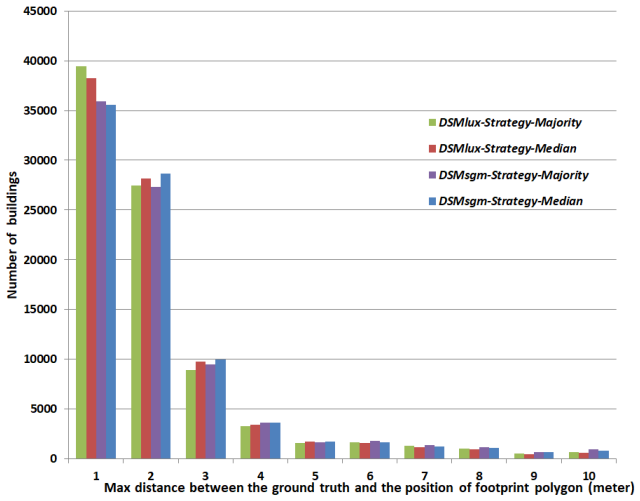


Fig. 4: Evaluation of the accuracy of footprint positions.

yielding 3D-LOD1 reconstruction model.

The key point in this method is that the RPC 'forward' model transforms directly world coordinates ($e.g$ lon, lat, z) to pixel coordinates (p, q). The RPC 'backward model' can transform (p, q, z) to (lon, lat, z), by using iterative method, since the polynomial for the backward model is not always provided by the vendor of satellite images.

3. EXPERIMENTS AND CONCLUSION

The proposed chain has been applied on a stereo pair of Pléiades satellite images at 50 cm spatial resolution over Marseille, France, with a theoretical elevation precision of 3.06 m. This precision is computed from the satellite azimuth and elevation for each image of the stereo pair. In our previous work [11], we have reported the accuracy of our rooftop extraction algorithm, yielding the mean intersection over union of 0.77, with 93% of detected buildings. Here we evaluate our chain by comparing the estimated footprint positions with the ground-truth footprints: 88146 building footprint contours have been manually drawn. This allows to evaluate both the proposed rooftop to footprint projection algorithm, and the quality of the built DSM, because the shift strongly depends on the estimated elevation.

Fig. 4 shows the accuracy in terms of distribution of distances between the automatically extracted building footprints and the corresponding ground-truth polygons: it describes the number of buildings where the position error $\epsilon < 1$ m, 1 m $< \epsilon <$ 2 m, *etc*. We have compared the use of DSM_{lux} and DSM_{sgm} for collecting elevation values, and two strategies to get the building elevation: by retaining the median or the majority value. Using DSM_{lux} with the majority and the median strategies yields 39407 (45%) and 38252 (43%) buildings with $\epsilon < 1$ m, respectively. The majority strategy performs better since it better estimates elevation values at building borders, while the median is influenced by higher slope values in the case of rooftops with a slope. The use of DSM_{lux} improves footprint positions when compared to DSM_{sgm}: using DSM_{lux} versus DSM_{sgm} with the majority strategy yields 39407 (45%) and 35928 (40%) buildings with $\epsilon < 1$ m, respectively, with the respective mean errors of 1.86 m versus 3.26 m. Fig. 2 illustrates visual quality of the reconstructed rooftops, footprints and DSMs. As can be observed, imprecisions in the DSM may significantly impact the final footprint results (e.g., a wrong position of the footprint at the top-center of Fig. 2c when using DSM_{sgm}), while with a good-quality DSM footprint positions and thus the building heights are correctly estimated.

In conclusion, we presented and validated an automatic pipeline for large-scale 3D-LOD1 reconstruction of buildings from satellite images. In particular, we proposed a robust solution for building footprint extraction, which allows to reconstruct in an efficient way footprints from the off-nadir image, using rooftop contours, DSM/DTM and RPC model. As future work we will extend the pipeline for 3D reconstruction in LOD2 (with non-flat roofs).

4. REFERENCES

[1] D. Poli and I. Caravaggi, "3D modeling of large urban areas with stereo VHR satellite imagery: lessons learned," *Natural Hazards*, vol. 68, no. 1, 2013.

[2] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE TPAMI*, vol. 30, no. 2, pp. 328–341, 2008.

[3] L. Duan and F. Lafarge, "Towards large-scale city reconstruction from satellites," in *ECCV*, 2016, pp. 89–104.

[4] B. Le Saux, N. Yokoya, R. Hansch, and M. Brown, "2019 ieee grss data fusion contest: large-scale semantic 3d reconstruction," *IEEE GRSM*, pp. 33–36, Dec 2019.

[5] S. Patil et al., "A new stereo benchmarking dataset for satellite images," *arXiv*, July 2019.

[6] H. Chen et al., "Multi-level fusion of the multi-receptive fields contextual networks and disparity network for pairwise semantic stereo," in *IGARSS*, 2019.

[7] R. Qin et al., "Pairwise stereo image disparity and semantics estimation with the combination of u-net and pyramid stereo matching network," in *IGARSS*, 2019.

[8] Z. Guo and Y. Xiuxiao, "On rpc model of satellite imagery," *Geo-spatial Information Science*, vol. 9, no. 4, pp. 285–292, Dec 2006.

[9] S. Tripodi et al., "Automated chain for large-scale 3d reconstruction ofurban scenes from satellite images," in *ISPRS PIA workshop*, 2019.

[10] B. Huang et al., "Large-scale semantic classification: outcome of the first year of inria aerial image labeling benchmark," in *IGARSS*, 2018.

[11] S. Tripodi et al., "Deep learning-based extraction of building contours for large-scale 3d urban reconstruction," in *SPIE remote sensing*, 2019.